DOCUMENT RESUME

ED 118 620 TM 005 110

AUTHOR Boyd, Joseph L., Jr.
TITLE The Effect of Format Change on Test Item Difficulty.
PUB DATE [Oct 74]
NOTE 4p.; Paper presented at the Annual Meeting of the Military Testing Association (16th, Oklahoma City, Oklahoma, October 21-25, 1974)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Multiple Choice Tests; *Test Construction; Testing Problems
IDENTIFIERS Item Difficulty (Tests)

ABSTRACT

This study evolved from a practical field situation that dictated an item format change. The item type in question is a variety of the multiple true-false item, widely used in state and municipal civil service examinations. Items are produced, not by combining pairs of independent true-false items, but as a means of salvaging not quite adequate four-choice multiple items. The original items may have no right answer: two, three, or four right answers; or one or two ambiguous or nonplausible responses. Items of this type were developed for use in a national testing program for automotive mechanics. After review by a number of test specialists and mechanics, it was decided that many items were faulty. A format variation was developed which seemed to be more clear cut, informal, and easier to read and understand. The revised items were used in a test battery. In practical terms, the overall effect of changing item format was to make the test items easier by an amount that would make mean percent correct scores higher by less than one percent. The real effect was to eliminate protests about the test questions. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED118620

TM005 110

THE EFFECT OF FORMAT CHANGE ON TEST ITEM DIFFICULTY

Joseph L. Boyd, Jr.
Educational Testing Service

There has been relatively little investigation of the effect that changing the format of multiple choice test items has on item difficulty. Such studies that have been done have dealt primarily with the effects of violating some of the ancient principles of item construction, or on the effect of using three, four, or five answer choices. Dunn, Goldstein, and Berkhouse (1956) studied the effects on difficulty, validity, and reliability of tests by violation of certain item writing principles: use of incomplete statement stems, avoidance of specific determiners, responses of equal length, and consistency of grammar between stem and responses. They found that items with specific determiners, with longer correct choices, and inconsistent grammar clues were easier, but no less reliable or valid, than items "correctly" written.

McMorris, Brown, Snyder, and Pruzek (1972) in a similar study, found that faults made the items easier, but that validity and reliability coefficients were virtually unchanged.

Board and Whitney (1972) found that violating good item writing practices generally made tests easier, as well as less reliable and valid. These, and similar studies, conform to the standard pattern of experimental design.

The study reported here evolved from a practical field situation that dictated an item format change. The item type in question is a variety of the multiple true-false item, widely used in state and municipal civil service examinations. Items are produced, not by combining pairs of independent true-false items, but as a means of salvaging not quite adequate four-choice multiple choice items. The original items may have no right answer; two, three, or four right answers; or one or two ambiguous or non-plausible responses. Conventionally, the stem and two viable responses are put into a format similar to the following:

Example 1. On a truck with a vacuum power booster, too much pedal pressure is needed to apply the brakes. What is a likely cause?

I. An air leak on the brake pedal side of the power booster cylinder

II. A leak in the diaphragm of the power brake booster

(A) I only (B) II only (C) either I or II (D) neither I nor II

Items in this format were developed for use in a national testing program for automobile mechanics. Raw four-choice items were written by instructors and service engineers who had had little item writing experience. Test specialists revised the items, which were then reviewed by separate panels of journeymen mechanics and service engineers. Many items were determined to be faulty, having no or multiple answers, or other ambiguities. The only feasible procedure was to recast the items into the format shown above.

After the items had been used in operational tests, there were protests from both mechanics and their employers that the items were too formal, too artificial, and too hard to read.

A format variation was developed, which seemed to be more clear cut, informal, and easier to read and understand. Item 2, below, is an example of this format.

Example 2. On a truck with a vacuum power booster, too much pedal pressure is needed to apply the brakes.

Mechanic A says that an air leak on the brake pedal side of the power booster cylinder could be the cause.

Mechanic B says that a leak in the diaphragm of the power brake booster could be the cause.

Who is right?
(A) A only (B) B only (C) both A and B (D) neither A nor B

The revised items were used in the test battery given in the Spring of 1974.

One hundred and four items, distributed among seven of the eight tests of the battery, were identified as having been used in both formats. The item difficulty levels were reviewed. Fifty-eight revised items were more difficult, 36 were easier, and 10 did not change in difficulty. Expressed in terms of percent correct, the mean changes in difficulty of the modified items were as follows:

Test	1	2	3	4	5	6	7
% Mean Change	-0.4	-0.1	0.0	+4.7	-1.3	-1.5	-0.1

For the 104 items, the mean difficulty level change was less than +.03%.

In practical terms, the overall effect of changing item format was

to make the test items easier by an amount that would make mean percent correct scores higher by less than one percent. The real effect of the format change has been to almost completely eliminate the protests about "those damn confusing, unreadable test questions."

REFERENCES

Board, C. and Whitney, D. R. "The Effect of Selected Poor Item-Writing Practices on Test Difficulty, Reliability and Validity." Journal of Educational Measurement, 1972, 9, 225-233

Dunn, T. F. et al. "Effect of Item Construction Principles on Difficulty, Reliability, and Validity." PRB Technical Research Note 64, November 1956, Personnel Research Branch, TAGO, DA.

McMorris, R. F. et al. "Effects of Violating Item Construction Principles." Journal of Educational Measurement, 1972, 9 287-295